Herpes simplex virus type 1 capsid protein, VP21, originates within the UL26 open reading frame

Stanley Person,* Sylvie Laquerre, Prashant Desai and John Hempel

Department of Molecular Genetics and Biochemistry, School of Medicine, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania 15261 U.S.A.

The goal of experiments reported here was to identify the genes that encode capsid proteins VP21 and VP24 of herpes simplex virus type 1 (HSV-1). Capsids were isolated from infected cells and the proteins were separated by SDS–PAGE. N-terminal amino acid sequence analysis of partial CNBr digestion products, and of intact VP21, showed that it is encoded within the UL26 open reading frame (ORF) of HSV-1 beginning with codon 248 and probably extending to the end of the ORF (codon 635). Similar analysis of digestion products confirmed that VP24 is specified by codons 1 to 247 at the 5' end of the UL26 ORF. Each of the seven known capsid proteins has now been assigned to an ORF.

The studies of Gibson & Roizman (1972), Heilman et al. (1979) and Cohen et al. (1980) identified seven capsid proteins in cells infected with herpes simplex viruses. They are designated, in order of decreasing Mr, VP5 (150K), VP19C (52K), VP21 (44K), VP22a (40K), VP23 (33K), VP24 (25K) and VP26 (12K). All except VP21 and VP24 are present at approximately 500 to 1500 copies per capsid. VP21 and VP24 are less abundant and are present at approximately 150 molecules per capsid (see Table 1).

Three types of capsids can be isolated from herpes simplex virus type 1 (HSV-1)-infected cells. They are visualized as light-scattering bands in sucrose gradients and are designated A, B and C, in order of increasing distance sedimented (Gibson & Roizman, 1972). They differ in protein and DNA composition and in their eventual fate in the infected cell. A and C capsids are similar in protein content, but only C capsids contain a genomic equivalent of DNA. VP22a is present only in B capsids and occupies the inner capsid space (Baker et al., 1990; Newcomb & Brown, 1989; Rixon et al., 1988).

VP22a is encoded by the UL26 open reading frame (ORF) (McGeoch et al., 1988) since a mutant virus, temperature-sensitive (ts) for the production of processed forms of VP22a (ts201), is rescued by restriction fragments located within UL26 (Preston et al., 1983). The ORF of UL26 encodes 635 amino acid residues in a single reading frame which are in a single reading frame and are 3'-coterminal (Holland et al., 1984; Liu & Roizman, 1991a). One transcript, designated UL26, is initiated 180 nucleotides upstream of the entire coding sequence and the second, more abundant transcript, designated UL26.5, is initiated near the centre of the UL26 ORF (1000 nucleotides downstream of the first transcript). The abundant polypeptides of the ICP35 family are encoded by this smaller transcript and are thought to be initiated at a methionine at codon 307 (Liu & Roizman, 1991a).

The entire UL26 ORF encodes a protein of approximately 80K that specifies a protease activity (Liu & Roizman, 1991b). In transient assays, plasmids that encoded this protein cleaved approximately 20 amino acids from the C terminus of the UL26.5 gene product. Furthermore, an identical self-cleavage was shown to occur in the protease (Liu & Roizman, 1991b; Preston et al., 1992). Gibson and his colleagues identified the
In order to pursue biochemical and genetic studies of capsid formation.

Products following expression of the UL26 ORF in protease may be found in Deckman mutation by deletion or insertional inactivation of coding sequences of the ORF. cge: Transient expression of UL26, codons 286 to 635, produced against cloned gene product precipitated the 12K protein from infected cells. Antibodies against VP26 precipitated the cloned gene product.

et al. product of cytomegalovirus strain Colburn (SCMV) N-terminal 249 amino acids of the homologous gene and size relative to VP5.

Table 1. Characteristics of B capsids

<table>
<thead>
<tr>
<th>Protein</th>
<th>Designation</th>
<th>Size ($M_r \times 10^3$)</th>
<th>Relative abundance</th>
<th>Reference*</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP5</td>
<td>UL19¶</td>
<td>149-1</td>
<td>54-0</td>
<td>aa: Davison et al. (1992)</td>
</tr>
<tr>
<td>VP19C</td>
<td>UL38</td>
<td>50-3</td>
<td>8-5</td>
<td>aa: Rixon et al. (1990)</td>
</tr>
<tr>
<td>VP21</td>
<td>UL26 (248–635)</td>
<td>38-0</td>
<td>2-1</td>
<td>aa: This publication</td>
</tr>
<tr>
<td>VP22a</td>
<td>UL26 (307–635)</td>
<td>32-4</td>
<td>18-1</td>
<td>cge: Liu &amp; Roizman (1991a)</td>
</tr>
<tr>
<td>VP23</td>
<td>UL18</td>
<td>34-3</td>
<td>9-3</td>
<td>nm: Desai et al. (1993)</td>
</tr>
<tr>
<td>VP24</td>
<td>UL26 (1–247)</td>
<td>24-3</td>
<td>1-3</td>
<td>aa: Davison et al. (1992) and this publication</td>
</tr>
<tr>
<td>VP26</td>
<td>UL35</td>
<td>12-1</td>
<td>6-7</td>
<td>aa: Davison et al. (1992)</td>
</tr>
</tbody>
</table>

¢ Abbreviations are as follows. aa: Amino acid sequence analysis of intact or CNBr-digested protein. mr: Marker rescue of ts mutation. nm: Null mutation by deletion or insertional inactivation of coding sequences of the ORF. cge: Transient expression of UL26, codons 286 to 635, produced against cloned gene product. The ICP35 family is indistinguishable by mobility and composition from the p40 family (which are precipitated by antibodies directed against a capsid protein thought to be VP22a). cg-ab: Cloned gene expressed in Escherichia coli. Polyclonal antibodies prepared against cloned gene product precipitated the 12K protein from infected cells. Antibodies against VP26 precipitated the cloned gene product.

¶ The range of $M_r$ values are those reported by Gibson & Roizman (1979), Heine et al. (1974), Davison et al. (1980) and DNA sequence (Davison & Scott, 1986).


¶ Map location of VP5 includes data for transcript mapping combined with antibody precipitation (Costa et al., 1984) and DNA sequence (Davison & Scott, 1986).

Cleavage sites and localized the protease activity to the N-terminal 249 amino acids of the homologous gene product of cytomegalovirus strain Colburn (SCMV) (Welch et al., 1991). Cleavage occurred between alanine and serine residues in sequences that are conserved in all of the known herpesvirus homologues of these proteins (Welch et al., 1991). For HSV-1 the cleavage sites are between residues 247 and 248, and at the C terminus, between residues 610 and 611. Dilanni et al. (1993) have confirmed these sites for HSV-1 by N-terminal amino acid sequence analysis of the autoproteolytic cleavage products following expression of the UL26 ORF in Escherichia coli. Further studies of the UL26-specified protease may be found in Deckman et al. (1992) and in Liu & Roizman (1993).

Davison et al. (1992) have shown that VP24, one of the two minor capsid proteins, is encoded at the 5' end of UL26. In similar experiments, we show that the remaining minor capsid protein, VP21, is encoded at the 3' end of UL26, and confirm that VP24 is encoded at the 5' end. These studies were undertaken to complete the assignment of each capsid protein to an ORF of HSV-1, in order to pursue biochemical and genetic studies of capsid formation.

Infected cells were harvested 48 h after infection and lysed (20 mM-Tris·HCl pH 7.5, 500 mM-NaCl, 1 mM-EDTA and 1% Triton-X100) to generate nuclei. Nuclei were pelleted, resuspended in 3 ml of lysis solution, and ruptured by sonication. The lysate was cleared by centrifugation, the supernatant layered onto the top of 24000 r.p.m. in an SW28 rotor (Beckman). Fractions containing B capsids were collected and the proteins were separated using SDS–PAGE (see Fig. 1). The seven capsid proteins are visible on the gel, as are additional proteins between VP5 and VP19C that commonly appear and may be contaminants. Intact VP21, VP22a and VP24 were prepared for amino acid sequence analysis using the Probe Design Kit according to the instructions of the manufacturer (Promega). Proteins were transferred to Immobilon–PVDF (polyvinylidene difluoride) membranes (Millipore), Coomassie blue-stained and sequenced directly (Matsudaira, 1987). To generate polypeptides for sequencing, VP21 and VP24 on Immobilon–PVDF membranes were subjected to partial digestion with CNBr (10 mg/ml in 70% formic acid for 15 h). They were eluted from the membrane, concentrated, and separated using tricine–SDS–PAGE (Schagger & von Jagow, 1987). Polypeptides were transferred to Problot membranes (Applied Biosystems), stained, and
Short communication

2271

VP5 - 97K
VP19C - 66K
VP21 - 45K
VP22a - 45K
VP23 - 31K
21-5K
VP24 - 21.5K
VP26 - 14.4K

Fig. 1. SDS-PAGE profiles of B capsids. Aliquots of B capsids were denatured by heating in the presence of SDS and 2-mercaptoethanol and loaded onto 15% polyacrylamide Laemmli gels with an N,N'-diallyltartardiamide cross-linker and the proteins were separated by electrophoresis. The gel was fixed and proteins were stained with Coomassie blue. The mobilities of the seven capsid proteins are marked on the left side of the figure and the $M_r$ markers are indicated on the right side.

the bands were excised for amino acid sequence analysis. Sequence analysis employed a Porton model 2090E gas–liquid phase amino acid sequencer with a diode array detector.

The data obtained after partial digestion with CNBr are shown in Fig. 2. CNBr cleaves after methionine residues, which occur at positions 1, 10, 221, 222, 253, 307, 482, 495, 505, 631 and 632 in the UL26 ORF of strain 17 of HSV-1 (McGeoch et al., 1988). VP21 was subjected to cleavage and the amino acid sequence determined for two polypeptides. The first polypeptide (approximately 14K in size) had the sequence YTPVAHYRPQVGE which was identical to the predicted amino acids at positions 506 to 518 of the UL26 ORF. The two polypeptides presumably result from non-quantitative cleavage at the methionines at residues 221 and 222. There is also a methionine at residue 10 of the UL26 ORF and analysis of an approximately 20K polypeptide, present in low quantity, revealed the sequence MLRDRwsLVAE. These amino acids are specified by codons 222 to 232 of the UL26 ORF. The two polypeptides presumably result from non-quantitative cleavage at the methionines at residues 221 and 222. There is also a methionine at residue 10 of the UL26 ORF and analysis of an approximately 20K polypeptide, present in low quantity, revealed the sequence EEPL corresponding to codons 11 to 14. These results are consistent with the recent assignment of VP24 to the 5' end of UL26 by Davison et al. (1992), and were obtained independently of their results.

To be certain that VP21 and VP22a were distinct molecules an attempt was made to determine the amino acid sequence of VP22a directly. Unlike VP21, its terminal residue was blocked which precluded its determination. Rixon et al. (1990) also reported that the N-terminal residue was blocked, as was the terminal residue of the homologous protein of SCMV (Gibson et al., 1990). A clone that included only residues 286 to 635, under the control of an exogenous promoter (Liu & Roizman, 1991a), expressed the four abundant members of the ICP35 family. Since the N-terminal amino acid of VP21 is 38 codons upstream of codon 286 the two proteins must be different. Schenk et al. (1991) also showed that the 45K capsid protein of SCMV, apparently a homologue to VP21, corresponds to an N-terminal amino acid extension of the more abundant VP22a homologue.

The UL26 ORF specifies three capsid proteins: VP21, VP22a and VP24. VP24 is encoded by the first 247 codons of UL26, and VP21 starts at codon 248 and probably extends to the end of the ORF (codon 635). Presumably VP24 and VP21 are generated by a protease
cleavage between alanine and serine at residues 247 and 248, respectively. The abundant polypeptides of the ICP35 family (and therefore VP22a) are specified by UL26 codons 307 to 635 (Liu & Roizman, 1991a). The C-terminal residue of the processed forms of VP21 and VP22a corresponds to codon 610.

The mutation in ts1201 (Preston et al., 1983) was mapped to a DNA fragment whose 3' terminus is now known to be located at codon 39 of UL26. Therefore, this mutation affects VP24 and shows it to be essential for viral replication. It would be surprising if VP22a was not essential for replication since it is intimately involved in DNA packaging. However, the less abundant VP21 specifies only 59 unique amino acids (residues 248 to 306), which are very poorly conserved in UL26 homologues (data not shown). This may indicate that it is not essential for virus growth. It is of interest to determine the role of all of the proteins in capsid formation and maturation. Since capsids were observed in the nuclei of cells infected with ts1201 at the non-permissive temperature (Rixon et al., 1988) it is possible that the three proteins are not required for empty capsid formation.

The assignment of the seven known capsid proteins to ORFs of HSV-1 is complete and is summarized in Table 1. The basis of the ORF assignments are also given in the table. Direct amino acid sequence analysis is now available for all of the capsid proteins except VP22a. It is interesting to note that among all of the HSV-1 proteins, VP5 is the most conserved and ranks slightly higher than the viral DNA polymerase (see for example, Telford et al., 1992). It is probably the most central protein from the point of view of capsid assembly and architecture.

We gratefully acknowledge the support and research facilities provided by Joseph C. Glorioso. This research was supported by Public Health Service grants from the National Institutes of Health. S. L. was supported by Le Fonds de la Recherche en Santé du Québec.

References


Liu, F. & Roizman, B. (1993). Characterization of the protease and


(Received 6 May 1993; Accepted 11 June 1993)